Claims

What is claimed is:

1.   A method of discovering permutation patterns from an input string having
5 a plurality of characters, each character being from an alphabet, the method comprising
the steps of:

selecting a new portion of the input string, the new portion differing from
a previously selected portion of the input string by at least one new character of the input
string;

10   determining one or more values for how many of the at least one new
characters are in the portion of the input string;

determining which, if any, names in a plurality of sets of names have
changed by selection of the new portion, the plurality of sets comprising a first set and a
plurality of additional sets, wherein the first set corresponds to all of the characters in the
15 alphabet and to values of how many of the characters of the alphabet are in the previously
selected portion, wherein the values are names for the first set, and wherein each
additional set comprises names corresponding to selected pairs of names from a single
other set; and

using changes in the names to determine the permutation patterns.
20

2.   The method of claim 1, further comprising the step of determining the
plurality of levels through the steps of:

determining the first set by determining values of how many of each of the
characters of the alphabet are in the previously selected portion; and

25   determining the additional sets by assigning names for a given additional
set to selected pairs of names from another of the sets, wherein each assigned name is
unique to the names for a selected pair.

3.   The method of claim 1, wherein the assigned names are codes.

YOR920030299US1   -15-

4.   The method of claim 3, wherein the codes are natural numbers.

5.   The method of claim 1, wherein the step of determining which, if any, names in a plurality of sets of names have changed determines that a name has changed and further comprises the step of determining that a new name is needed for the changed name.

6.   The method of claim 5, wherein the step of determining which, if any, names in a plurality of sets of names have changed further comprises the step of selecting a new name, not currently in use in the sets of names, for the changed name.

7.   The method of claim 1, further comprising the step of determining for a name that has changed in the sets of names, a location in the input string that corresponds to the changed name.

8.   The method of claim 7, wherein the changed name corresponds to at least two characters of the input string and a location in the input string of a given character of the at least two characters is chosen as the determined location.

9.   The method of claim 1, wherein each of the names in the sets of names corresponds to a pattern, and wherein the step of using changes further comprises the step of selecting permutation patterns from the patterns.

10.   The method of claim 1, further comprising the step of comparing names that have changed in the sets of names to a database comprising a plurality of stored names.

11.   The method of claim 1, wherein the additional sets have names corresponding to only a single pair of names from another set.

12.   The method of claim 1, wherein the step of using changes further
5 comprises the step of correlating the changed names with permutation patterns.

13.   The method of claim 12, wherein the step of determining which, if any, names in a plurality of sets of names further comprises, for each changed name, updating a count corresponding to that changed name, and wherein the method further comprises
10 the step of:
           performing the steps of selecting, determining one or more values, and determining which, if any, names in a plurality of sets of names until the entire input string has been selected.

15 14.   The method of claim 13, wherein portions selected have a predetermined size, and wherein the method further comprises the step of selecting a number of predetermined sizes and performing the steps of selecting, determining one or more values, and determining which, if any, names in a plurality of sets of names for each of the predetermined sizes.
20
15.   The method of claim 14, wherein the step of using changes further comprises the step of determining permutation patterns corresponding to counts greater than or equal to a predetermined count.

25 16.   The method of claim 15, further comprising the step of determining maximal permutation patterns from the determined permutation patterns.

17.   The method of claim 16, wherein the step of determining which, if any, names in a plurality of sets of names further comprises the step of determining location

YOR920030299US1   -17-

lists for each of the names corresponding to permutation patterns, and wherein the step of determining maximal permutation patterns further comprises the steps of comparing location lists for permutation patterns and eliminating duplicate permutation patterns by using the location lists.

18.   The method of claim 1, wherein the at least one character is a single character and wherein the step of selecting further comprising selecting a portion of the input string that differs from the previously selected portion of the input string by moving a window one character, from the previously selected portion, along the input string, the window selecting the new portion of the input string.

19.   The method of claim 1, wherein the sets of names are stored in a balanced search tree.

20.   An apparatus for discovering permutation patterns from an input string having a plurality of characters, each character being from an alphabet, the apparatus comprising:

a memory;

at least one processor coupled to the memory, the at least one processor configured:

to select a new portion of the input string, the new portion differing from a previously selected portion of the input string by at least one new character of the input string;

to determine one or more values for how many of the at least one new characters are in the portion of the input string;

to determine which, if any, names in a plurality of sets of names have changed by selection of the new portion, the plurality of sets comprising a first set and a plurality of additional sets, wherein the first set corresponds to all of the characters in the alphabet and to values of how

many of the characters of the alphabet are in the previously selected portion, wherein the values are names for the first set, and wherein each additional set comprises names corresponding to selected pairs of names from a single other set; and

5   to use changes in the names to determine the permutation patterns.

21.   The apparatus of claim 20, wherein the at least one processor is further configured, in order to determine the plurality of levels:

10   to determine the first set by determining values of how many of each of the characters of the alphabet are in the previously selected portion; and

to determine the additional sets by assigning names for a given additional set to selected pairs of names from another of the sets, wherein each assigned name is unique to the names for a selected pair.

15

22.   The apparatus of claim 20, wherein the at least one processor is further configured, when determining which, if any, names in a plurality of sets of names have changed determines that a name has changed to determine that a new name is needed for the changed name.

20

23.   The apparatus of claim 20, wherein the at least one processor is further configured to determine, for a name that has changed in the sets of names, a location in the input string that corresponds to the changed name.

25 24.   The apparatus of claim 20, wherein each of the names in the sets of names corresponds to a pattern, and wherein the at least one processor is further configured, when using changes in the names, to select permutation patterns from the patterns.

25.   The apparatus of claim 20, wherein the additional sets have names corresponding to only a single pair of names from another set.

26.   The apparatus of claim 20, wherein the at least one processor is further configured, when using changes in the names to determine permutation patterns, to correlate the changed names with permutation patterns.

27.   The apparatus of claim 20, wherein the at least one character is a single character and wherein the at least one processor is further configured, when selecting a new portion of the input string, to select a portion of the input string that differs from the previously selected portion of the input string by moving a window one character, from the previously selected portion, along the input string, the window selecting the new portion of the input string.

28.   The apparatus of claim 20, wherein the sets of names are stored in a balanced search tree.

29.   An article of manufacture for discovering permutation patterns from an input string having a plurality of characters, each character being from an alphabet, the article of manufacture comprising:

a computer readable medium containing one or more programs which when executed implement the steps of:

selecting a new portion of the input string, the new portion differing from a previously selected portion of the input string by at least one new character of the input string;

determining one or more values for how many of the at least one new characters are in the portion of the input string;

determining which, if any, names in a plurality of sets of names have changed by selection of the new portion, the plurality of sets

comprising a first set and a plurality of additional sets, wherein the first set corresponds to all of the characters in the alphabet and to values of how many of the characters of the alphabet are in the previously selected portion, wherein the values are names for the first set, and wherein each additional set comprises names corresponding to selected pairs of names from a single other set; and

using changes in the names to determine the permutation patterns.

5

10